Check for updates

TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 17 (2025) 662–680
© 2025 The Author(s). *Topics in Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society.
ISSN: 1756-8765 online
DOI: 10.1111/tops.70001

This article is part of the topic "2023 Rumelhart Prize Issue Honoring Nick Chater," Morten H. Christiansen and Mike Oaksford (Topic Editors).

The Limited Place in Cognitive Space

Carl J. Hodgetts,[a] Ulrike Hahn[b]

[a]*Department of Psychology, Royal Holloway, University of London*
[b]*Department of Psychological Sciences, Birkbeck, University of London Malet Street, Bloomsbury London WC1E 7HX*

Received 28 February 2024; received in revised form 6 February 2025; accepted 11 February 2025

Abstract

How we judge the similarity between stimuli in the world is connected ultimately to how we represent them. Because of this, decisions about how we model similarity, either in terms of human behavior or patterns of neural activity, can provide key insights into how representations are structured and organized. Despite this, psychology and cognitive neuroscience continue to be dominated by a narrow range of similarity models, particularly those that characterize similarity as distance within "cognitive space." Despite the appeal of such models, their topological nature places fundamental constraints on their ability to capture relationships between objects and events in the world. To probe this, we created a stimulus set in which the predicted similarity relationships (based on an alternative model of similarity) could not be simply embedded within Euclidean space. This approach revealed that the spatial model distorts these predictions, and the perceived similarities of human observers. These findings indicate that cognitive spaces—that underlie much recent work probing both visual and conceptual representations in cognitive neuroscience—are limited in fundamental ways that restrict their theoretical and practical utility.

Keywords: Similarity; Mental representation; Conceptual knowledge; Multidimensional scaling; Vision; Neural coding; Object perception

Correspondence should be sent to Ulrike Hahn, Department of Psychological Sciences, Birkbeck, University of London, Malet Street, Bloomsbury, London WC1E 7HX, UK. E-mail: u.hahn@bbk.ac.uk
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Imagine encountering a shrub with fruit hanging from its branches. Are these the same tasty and nutritious fruit you ate several months ago? The fruits seem the same color, and the leaves seem a similar shape, but the fruits appear to be sitting upon the leaves, rather than hanging below, as you remember. Are the fruits in front of you still safe to eat? Simple examples such as this illustrate the fundamental, and adaptive, role of similarity in memory, categorization, and the generalization of behaviors.

How we judge the similarity between stimuli in the world is connected, ultimately, to how we represent them (Edelman, 1998; Hahn, 2014; Roads & Love, 2024). In other words, how different the two shrubs in our example seem depends on what information you have encoded. Does your memory representation include not only the color of the fruit, and the shape of the leaves, but the spatial relationship between these properties? Depending on the answer to this question, these two shrubs may look quite different, or completely the same.

How one chooses to model human similarity, therefore, has significant implications not only for capturing human behavior across a diverse range of contexts, but for our broader understanding of how human mental representations are structured and organized. Despite this, research across many related disciplines—spanning human and animal psychology, cognitive neuroscience, and machine learning—continues to be dominated by a small subset of models, particularly spatial models of similarity (Shepard, 1957). Under such models, similarity is broadly conceptualized as distance within a multidimensional cognitive space. An object's position within the space is determined by its values along a set of relevant psychological dimensions. Proximities between objects or exemplars within this metric cognitive space, therefore, can be considered to reflect similarity, whereby nearby items are highly similar, and distant items are highly dissimilar, as determined by the specific distance metric applied (e.g., Euclidean distance). This theoretical framework has provided the basis for several highly influential models of human recognition within cognitive science (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986; Stewart & Brown, 2005), and continues to gain popularity within contemporary cognitive neuroscience, particualrly given evidence that the neural mechanisms for representing physical space may also encode conceptual space (Bellmund, Gärdenfors, Moser, & Doeller, 2018; Bottini & Doeller, 2020; Constantinescu, O'Reilly, & Behrens, 2016; Mur et al., 2013).

Notably, many of the studies that have applied such spatial models to assess human recognition and conceptual knowledge—both at the behavioral (Hahn, Bailey, & Elvin, 2005) and neural levels (Bellmund et al., 2018; Constantinescu et al., 2016)—have used stimuli that are well-suited to those models in the first place, such as stimuli that vary on a few continuous dimensions. As spatial models operate upon relatively simple object representations (i.e., locations within an *n*-dimensional space), the simple and artificial stimuli often used will not pose serious problems for such models (or individual studies) and thus provide limited insight into the nature or complexity of the underlying object representations—or indeed their neural substrates. For instance, many naturalistic stimuli outside the domain of colors and sounds cannot be easily represented as sets of continuous feature dimensions (Tversky & Hutchinson, 1986). Not only are many naturalistic objects characterized by a large number of such dimensions, but, as noted above, these dimensions must be arranged in a particular way to be appropriately categorized, discriminated, and so on (Biederman, 1987). In other

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

words, as soon as representations become structured, knowledge-based, or analogically rich, these basic similarity models appear fundamentally limited (Hahn, Chater, & Richardson, 2003; Hodgetts, Hahn, & Chater, 2009; Leech, Mareschal, & Cooper, 2007). Moreover, Euclidean models—and by extension cognitive spaces—place an upper bound on the number of points that can share the same "nearest neighbor" (Odlyzko & Sloane, 1979; Tversky & Hutchinson, 1986). Within a two-dimensional space, for instance, a maximum of six distinct items may share the same nearest neighbor, corresponding to the center and six vertices of a regular hexagon. As in the case of structured representation, this may pose few problems for the highly specific and simple stimuli used in many studies, but may lead to systematic distortions within specific stimulus domains, such as those which are hierarchical in nature (e.g., conceptual structures; Tversky & Hutchinson, 1986).

1. Trouble in the neighborhood

The first aim of this paper is to use these basic topological characteristics to demonstrate the limitations of spatial models. Let us first consider the set of items in Fig. 1a. These items were generated from a stimulus domain that has been used previously to evaluate structural models of similarity (Cheries, Newman, Santos, & Scholl, 2006; Hodgetts et al., 2009, 2023; Hodgetts & Hahn, 2012; Larkey & Markman, 2005), allowing us to directly compare spatial models with alternative accounts of similarity—in particular the transformational model of similarity (Hahn et al., 2003; Hodgetts et al., 2009). Critically, the transformational model not only assumes the representation of structural information within this domain (Hodgetts et al., 2009), but also gives rise to fundamentally different topologies for the similarity space than do spatial models (outlined in Fig. 1b–g). The items themselves involve pairs of simple geometric shapes, and these can be described by a basic code where each letter denotes a unique shape (e.g., the comparison "square to the left of triangle vs. triangle to the left of square" can be referred to as *AB/BA*). Under a transformational model, similarities between all possible comparisons can be described via sequences of psychological operators (whereby longer sequences reflect more complex transformations and, in turn, lower perceived similarity). Three psychological operations have been proposed ("create," "apply," and "swap"), which have been shown to accurately capture similarity judgments and speeded same-different judgments within this domain (Hodgetts et al., 2009, 2023; Hodgetts & Hahn, 2012; see Methods).

Fig. 1b–f provides an example of how, even for this simple stimulus set, it is possible (under this structural model) to derive neighborhood relations in that are incompatible with a purely spatial representation (see Fig. S1 for a further example). Under the assumed similarity relationships, there are four items equally similar to the hypothetical item *AA* (Fig. 1b). However, there are no four spatial locations that would allow these items to be equidistant to *AA* while also respecting the similarity relations among them. While it is no problem to embed three of these items (see Fig. 2c,d), there is no possible location for the fourth item (item *CA*) that respects all relative distances predicted by the transformational model (Fig. 1e–g). Crucially, there is nothing complicated about the stimuli, or mysterious about the pairwise transformation distances—both of which are straightforwardly defined (Hodgetts et al., 2009,

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

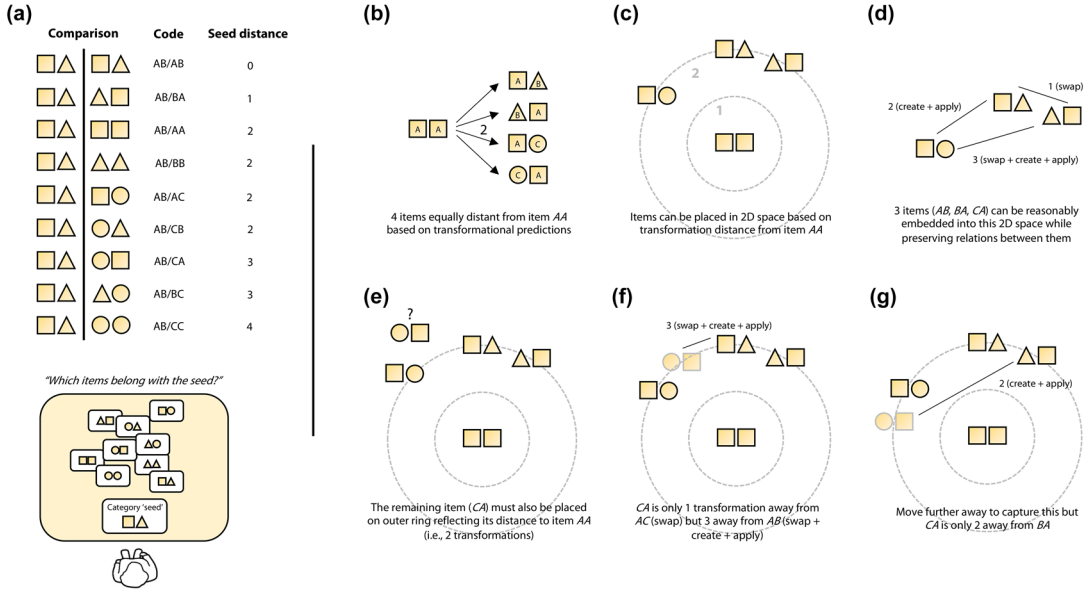

Fig. 1. The item set and a naturalistic "free sorting" task used to probe neighborhood constraints in spatial models. (a) Comparisons between a target item AB and nine comparison stimuli. Each item can be described by a code where each feature on the shape dimension can be denoted by a unique letter (upper). The transformational model prediction for each comparison, derived from the coding scheme (see Methods), is shown. Similarity between such items may be assessed using a free sorting paradigm (lower). Such a task begins with a "seed" item (item AB) and participants must select items that they feel belong with the item. (b–g) The placement of items within a two-dimensional Euclidean spatial representation based on distances derived from an alternative transformational model of similarity (see main text).

2023; Hodgetts & Hahn, 2012)—it is simply that the more general notion of transformation distance (of which Euclidean or spatial distance is a special case) supports relationships that cannot be embedded within the restrictive topology of spatial distance.

Imagine such similarities being operational within a simple similarity-based task, such as unsupervised categorization (Regehr & Brooks, 1995). In such a task, participants are given a specific "seed" item (see Fig. 1a) and must decide which items from the broader stimulus set "belong" with this seed item. This type of "free sorting" task is used widely because of its close correspondence to category construction in the real world (Milton & Wills, 2004), and similarity has been shown to be extremely important in determining how items are classified in such a task (Close, Hahn, Hodgetts, & Pothos, 2010; Love, Medin, & Gureckis, 2004; Love & Gureckis, 2002). Past research leads one to expect that the probability of inclusion in the category should decrease as a function of distance to the seed. To show the basic incompatibility between the spatial and transformational model in this domain, one can take the pairwise transformation distances and use multidimensional scaling (MDS) to derive a spatial solution for them (see Fig. 2a, upper). The topological tension of this solution is reflected in a resulting stress value of 0.19, which is considered an unacceptable, or "imprecise," level

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

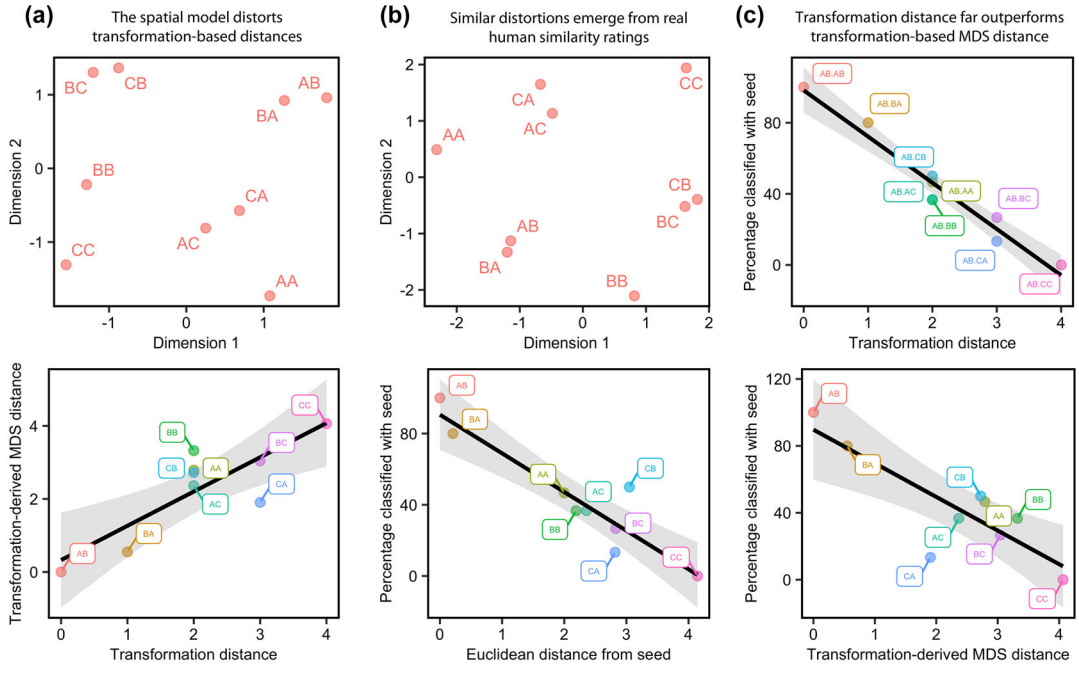

Fig. 2. Evidencing nearest neighbor constraints for our item set and task. (a) The MDS representation of the pairwise transformational predictions for this item set (upper), and the relationship between these distances (labeled, "Transformation-derived MDS distances") and the original predictions (lower). The lower plot shows how MDS distorts the original similarity predictions (see main text). (b) The MDS representation of the human pairwise similarity ratings for all items in the stimulus set (upper), and how these relate to classification performance in the free-sorting task (lower). (c) The relationship between the transformational model and classification performance (upper). This is presented above a plot showing the association between the transformation-derived MDS distances and classification performance. Nine data points are shown on each scatter plot reflecting each item comparison in the task. The 95% confidence interval is also shown.

of data distortion (Kruskal, 1964). This tension is also apparent in the annular configuration that is produced, which is commonly seen when fits are poor and the underlying configuration is nonannular but incompatible (Goodhill, Simmen, & Willshaw, 1995; Fig. 2a, upper). Moreover, a correlation plot between these transformation-derived MDS distances and the original transformational distances (Fig. 2a, lower) shows how specific individual items seemingly "break" the spatial model, in particular items *BB* and *CA* (Fig. 1; see also Fig. S1).

To demonstrate the psychological relevance of this, we gave this task to 30 participants. This allowed us to confirm the difficulties of the spatial model in capturing the similarity relationships between items in this stimulus domain—both in terms of pairwise similarity judgments and classification behavior in the sorting task. To derive spatial model predictions for items in the set, we applied MDS to pairwise similarity ratings from an independent set of raters (see Methods). The distances derived from the MDS representation provided a strong fit of the classification data ($r(7) = -.94$, $p < .001$, $R^2 = .88$; see Fig. 2b), yet there are indi-

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

cations that certain item relationships are not easily embedded within this Euclidean solution. First, a fair-to-poor degree of stress is observed for this MDS solution (Stress-1 = 0.11). And further, the items in this ratings-derived MDS representation also form an annular configuration (Fig. 2b), which, under closer inspection, has obvious similarities to the one derived from the transformational predictions. Finally, while the spatial model provides a good fit of the classification data, the transformational model's fit is near-perfect ($r(7) = -.98$, $p < .0001$), accounting for 95% of the variance (Fig. 2c). A likelihood ratio based on these fits suggests that the data are 51 times more likely to occur under a transformational model (see Glover & Dixon, 2004).

Two final considerations further confirm the topological difficulties of the spatial model. First, we used the transformation-derived MDS distances (depicted in Fig. 2a) to predict classification behavior in our human participants. Although statistically significant ($r(7) = -.84$, $p = .004$, $R^2 = .71$), the fit is less accurate than that obtained from original transformational model predictions, implying that MDS distorts these predicted similarities (Fig. 2c). This is confirmed by an increased Stress-1 value of 0.19. Second, the residuals that were problematic when deriving an MDS representation of the transformation distances (see Fig. 2a) are similarly problematic when those distances are derived from human similarity judgments (namely, items *BB* and *CA*). Notably, the original transformational predictions fit these comparisons very well (see Fig. 2c, upper), suggesting that these poorer fits relate to the spatial model in general and not, for example, the sample size from which the pairwise ratings were derived.

Multidimensional cognitive spaces, in which similarity is determined by Euclidean distances between objects, are central to many influential models of human cognition (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986; Stewart & Brown, 2005), and have recently been proposed as the "primary representational format" for information coding in the brain (Bellmund et al., 2018). However, are such models even appropriate for estimating similarity between simple stimuli when they are not amenable to those models in the first place? Our data show how a stimulus domain, which has been used successfully to provide evidence for transformational models of similarity in the context of similarity ratings (Hodgetts et al., 2009, 2023) and speeded same-different judgments (Hodgetts & Hahn, 2012), can be used to create similarity relationships that cannot be simply embedded within a Euclidean metric space (Fig. 1b–g). Indeed, one of the fundamental limitations of so-called "spatial codes" is that they place an upper bound on the number of distinct items that can be equally similar to any other (Tversky & Hutchinson, 1986). While this may pose few problems within specific and artificial stimulus domains that are designed specifically with Euclidean spaces in mind (e.g., manipulating continuously valued dimensions), it places a fundamental limit on such representations as a generalized model of human cognition, given that multiple, complex object and events must be co-represented, remembered, and discriminated.

Our approach above revealed that spatial models can systematically distort the relationship between objects, even within a relatively simple stimulus domain. Such distortions were evident when spatial/MDS representations were derived from human similarity ratings (and then applied to novel classification data in a free-sorting task) and reflected those seen when spatial representations were derived directly from the predictions of the transformational model.

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This analysis demonstrated how certain item similarities were simply not compatible, leading to a lack of precision in model fits—even within this relatively "simple" stimulus domain.

2. The fundamental incompatibility

The violation of the nearest neighbor constraint, however, is itself only one small manifestation of a more fundamental mismatch between the above transformation "space" and an MDS-style spatial representation. To demonstrate this, think of a stimulus whose construction is determined by four separable binary "features" (dimensions), and an item set that exhausts all possible combinations, giving rise to 16 items. To represent these as distinct items in a spatial framework, one needs four dimensions (one for each feature), giving rise to a four-dimensional hypercube. Distance between these items is then (given a City Block metric) the sum of the dimensional differences (or for a Euclidean distance, the square root of the sum of the squared differences; see Shepard, 1957).

One can think of these "moves" along the individual dimensions as four simple "transformations" and, hence spatial models as transformational models with restrictive transformations (Hahn et al., 2003). But a transformation set like the one described above creates a very different transformation space with a very different dimensional structure, given that there are only three possible transformations. This gives rise to fundamental incompatibilities between the resultant distribution of distances, even within this higher-dimensional space.

To illustrate this, we constructed an exhaustive 16-item set that is a variation of the items presented in our free-sorting task above. Each item in the set is representable as a 4-bit vector that represents the following stimulus aspects: triangle/square and light/dark for the left-hand component, and triangle/square and light/dark for the right-hand component, yielding the set depicted in Fig. 3a. Applying our transformation scheme from above ("create," "apply," and "swap") to those items gives rise to the proximity matrix of pairwise distances shown in Fig. 3b. We can contrast that with the proximity matrix that arises from the 16-corner hypercube based on the same 4-bit vectors (Fig. 3c). A key feature of this second matrix is that the *distribution* of distances to all remaining items is *the same* for each item (Fig. 4d; see Supplementary Text and Fig. S2 for a detailed explanation).

Of course, when considering this hypercube proximity matrix (Fig. 3c), we would not be expecting it to match perceived similarity exactly, because different dimensions or features may carry different weights (capturable by shrinking or stretching the hypercube along a given dimension; see Nosofsky, 1991). But that is necessarily a *systematic* operation, that will affect the whole set of items. Its effect can be visualized by considering the dendrograms at the sides of the heatmap. Each branching corresponds to one of the four binary features' dimensions. Altering the relative weight of a feature is equivalent to stretching or shrinking the length of all sub-branches that reflect that contrast. But this means that the effects must simultaneously act on all items in the entire set, which, in turn, means specifically, that the distance *distributions* for all items remain equal (though the individual distances have changed; see Supplementary Text and Fig. S2 for a visualization of this). Most importantly, there is no combination of such dimension-wide operations that could make it the case that, selectively, one item has no

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

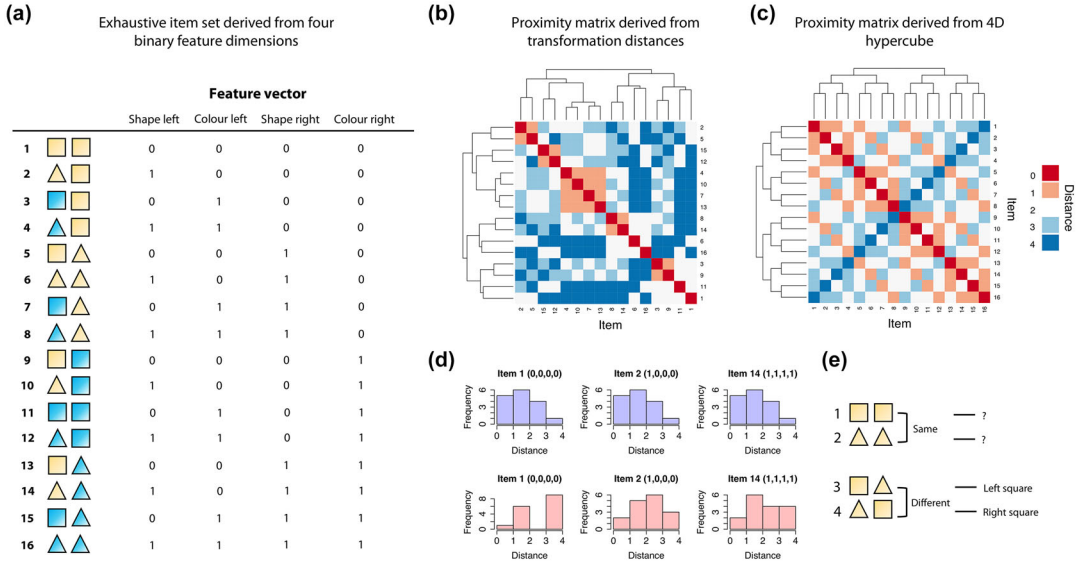

Fig. 3. Exploring the topological limitations of high-dimensional spaces. (a) A 16-item set where each item is determined by a vector of four binary features: square/triangle and yellow/blue for the left-hand component, and square/triangle and yellow/blue for the right-hand component. (b) A 16×16 proximity matrix derived from the transformational model predictions. (c) A 16×16 proximity matrix derived from the four-dimensional hypercube. The color code for both matrices is shown on the right. (d) Histograms showing the distribution of distances for the hypercube-transformation-based models for three example items (items 1, 2, and 14). Plots demonstrate that the distribution of distances to all remaining items is identical for each item under a hypercube dimensional structure. (e) "Relational features" (e.g., "triangle left/triangle right") may be plausible for specific items/comparisons (e.g., items 3 and 4), but cannot be assigned values for others (e.g., items 1 and 2; see main text).

nearby neighbors. By definition, any dimensional shrinking or stretching will affect all items, jointly impacting entire subsets of items in the same way.

3. Stimulus bias and asymmetry

The only way to get differences beyond that is to appeal, formally and intuitively, to notions of individual stimulus bias. This is already necessary to make spatial models compatible with another fundamental property of transformation distance (and human similarity data), namely that it can be asymmetrical (Hahn, Close, & Graf, 2009; Hodgetts & Hahn, 2012; Op De Beeck, Wagemans, & Vogels, 2003). This problematic notion of asymmetric similarity has already been invoked against spatial models (Tversky, 1977), and it has since been proposed that asymmetries can be captured in spatial models through the addition of a stimulus bias term (Nosofsky, 1991). The introduction of such a bias term not only makes the spatial model formally identical to an additive feature model, but also allows the spatial model to circumvent the dimension-dependent nearest neighbor constraint discussed above (Nosofsky, 1991). This does not, however, provide an adequate solution here.

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Fundamentally, "stimulus bias" is a property of individual items that operates *in addition* to a symmetric notion of similarity. It thus requires independent motivation and evidential support. Addressing this, a previous study tested the asymmetric predictions of the transformational model using a set of items like those presented here (Hodgetts & Hahn, 2012). Not only were the expected directional differences observed (i.e., participants were *faster* to tell two items apart when the comparison was in the direction of a "long" transformation), but the data also showed that intrinsic stimulus properties, such as complexity or symmetry, could not account for the observed patterns. Empirical tests of bias models are possible because even this machinery, which focuses on the properties of individual items, has wider systematic consequences: unlike dimensional shrinking or stretching, adding individual bias selectively affects only a single row. But there, its effects across the entire row must be considered, and those include the impact of the bias on *self-similarity* (Bailey & Hahn, 2005; Krumhansl, 1978; Nosofsky, 1991). In Hodgetts and Hahn's data, those self-similarity predictions were disconfirmed.

The ultimate lesson, here, is always the same: positing "features," "dimensions," or individual stimulus properties such as "salience" may seem locally plausible when considering just a few comparisons. But it then conceptually fails when the item set under consideration is expanded (see also Hodgetts, Close, & Hahn, 2023). For example, it may seem trivial to capture the current stimuli in terms of basic features (square/triangle; yellow/blue) supplemented with "features" that recast the relevant relations (same/different; triangle-on-left/triangle-on-right). That this is ultimately not feasible is documented across many studies of similarity (Goldstone, Medin, & Gentner, 1991; Hahn, 2014; Hodgetts et al., 2023; Markman & Gentner, 1993). And the reason it is not, is that—while locally plausible—such "relational features" will invariably run into comparisons where the scheme breaks down as shown in Fig. 4e: While such "relational features" may seem plausible for items 3 and 4, they cannot even be assigned values for items 1 and 2 (in the case of 1, because there are no triangles, so no value of the "triangle left/triangle right" feature applies, and the case of object 2, because both values of the binary relation apply simultaneously). Relations cannot, at scale, be reduced to features. So, the fact that a spatial model is formally equivalent to an additive feature model under some constraints is not a way out: neither are adequate in this context.

4. From low- to high-dimensional spaces

The only route "out," here, for both spatial *and* featural models, would be to allow essentially arbitrary dimensions or features. For example, in the case of bias-supplemented spatial models, it was previously noted:

> "The upshot of this analysis is that for stimuli that are representable as vectors of 'on-off' features, a bias-supplemented MDS model can always be formulated that is formally identical to the additive version of Tversky's (1977) feature-contrast model. Note that although such a formulation may require a high-dimensional solution (with the dimen-

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

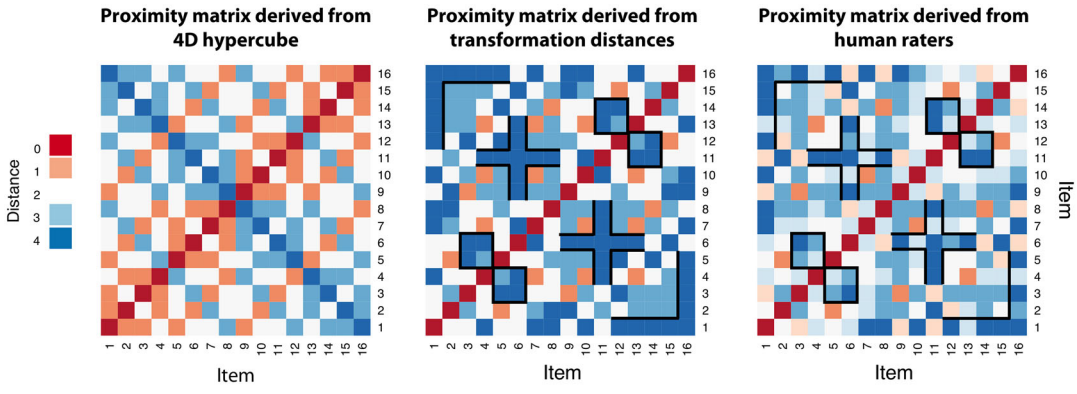

Fig. 4. Evidencing nearest neighbor constraints for our item set and task. Human proximity data from a pairwise rating task using the stimulus set shown in Fig. 3 are compared to the proximity matrix that follows from a hypercube representation and from the transformation distances, respectively. The two target matrices, hypercube distance and transformation distance, are repeated from Fig. 3 but without clustering dendrograms so that all rows and columns in the three plots are ordered by stimulus number. Black lines highlight regions of visual correspondence. Statistical analysis of correspondence can be found in the text.

> sions being binary-valued), there is no logical necessity that MDS solutions consist of only a few continuous dimensions."—Nosofsky (1991, pp. 104–105)

This may seem innocuous (and there may be contexts where it genuinely is), but, in general, it is not. Increasing the dimensionality of the set is arguably what one is intuitively appealing to when one says the stimuli give rise to "higher order relations" such as same/different or "left/right") and one can, of course, represent our 16-item set in an even higher dimensional space (in fact, there are infinitely many such spaces). Likewise, the mere fact that we can draw the dendrograms at the sides of Fig. 3c indicates that there is some weighted "feature set" that will reflect those distances. This route will *always* be possible. However, it adds nothing and achieves nothing when the only thing those representations do is directly implement the numerical values of the proximity matrix, because this adds nothing to the proximity matrix itself. Crucially, the dimensions of that higher-dimensional set *now no longer correspond to the psychologically salient factors* (captured in the 4-bit vector of Fig. 3a) that went into constructing it—as holds by necessity, precisely because left/right (as a relation) is not a *feature* on par with shape or color the first place. But if the dimensions of the newly derived even higher dimensional hypercube no longer have obvious psychological interpretation, the representation has little value.

The mere fact that a set of numbers can be represented in an entirely general, abstract mathematical framework holds no more intrinsic interest than does the fact that the same set of numbers can be represented as a table, a list, a vector, or a pie chart. What matters from the perspective of cognitive science is that representation is *informative* both with respect to explanation and prediction. In the context of cognition, what matters is explanation and prediction of how those numbers come about.

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The whole point of featural or spatial models of similarity was that their features or dimensions corresponded to psychologically (or potentially neurally) meaningful aspects of the stimulus. Where they no longer do so, they hold no methodological or conceptual value. If all they do is re-represent a proximity matrix, without adding information on how the values of that matrix may be derived, then they simply contribute nothing as analytical frameworks. One is just as well-off using the proximity data itself.

Very finally, nothing shown here hinges particularly on the transformational account of similarity, let alone the specific transformations chosen ("create," "apply," "swap"). These are just intuitive renditions of what it would mean to make two structured representations identical (swap arguments for a predicate, replace value for a predicate, replace predicate; see Gentner, 1989). Given the intrinsic conceptual link between similarity and distance (as degree of departure from identity), we suspect that any meaningful theory of similarity can be given a transformational explication, without that necessarily implying a process-level theory whereby similarity is computed (where it is) via a search for transformation. In other words, the transformation distances between stimuli here are a reflection of their underlying representational structure. Hence, the driving force, here, is that relations are not, and can never adequately be, "features." And the reason more general, empirically adequate "transformation spaces" cannot be projected into constrained hypercube models is that the limited "transformations" of the latter (i.e., "move along a dimension") are, by definition, available at each step and, by definition, independent: it is always possible to move by one step along any of the *n*-dimensions of a hypercube, and the order in which steps are taken has no impact on the ultimate point reached. The transformations in our simple scheme, by contrast, are not like that: one can only "apply" what already exists, and one can only "swap" where differences exist: no swap is possible for item 1 in 3E, and nothing can be "applied" somewhere it does not already exist (which is why this item has no neighbors one transformation away). This failure of independence is not restricted to our particular transformation scheme, but characterizes also a much broader set of visual transformations that play a role in perception such as scaling, dilation, rotation, or reflection (Graf, 2006; Lawson, 1999). These operations need not commute, meaning the order in which they are applied matters. This gives rise to a fundamental incompatibility between more general transformation "spaces" and the restrictive properties of MDS-style spatial models. This extends to any account that wishes to claim that mental representation in locally Euclidean manifolds is psychologically or neurally plausible, including recent models of concept learning and conceptual structure based on deep neural networks (Sorscher, Ganguli, & Sompolinsky, 2022). To what extent these limitations are overcome by other nonstandard spatial models (e.g., Nickel & Kiela, 2017) demands further attention. Likewise, our results provide added motivation for research, for example, on how deep learning methods can deal with graph representations (Li, Huang, & Zitnik, 2022) as a known means for representing representational structure. Finally, they highlight the need to understand in detail how, and to what extent, the increased representational abilities of transformer-based models (e.g., their ability to implement [nonclassical] variable binding; Buckner, 2023; Millière, 2024) meet the demands of adequately representing structure in nonlinguistic domains.

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5. Demonstrating the psychological relevance

The only step that remains is to show that, indeed, the preceding theoretical considerations have relevance to actual empirical data from human participants. Addressing this, we asked 240 human participants to rate *all* pairwise similarities within the stimulus domain laid out in Fig. 3a (see Methods) and correlated the resulting proximity matrix with those derived from both the 4D hypercube and the transformational model (see Fig. 4). Notably, the transformation-based matrix is highly correlated with the human perceived similarities ($r(240) = .82$, $R^2 = .67$), whereas the predicted relationships from the 4D hypercube provide a much poorer fit of the human data ($r(240) = .46$, $R^2 = .21$). Not only can this difference be observed by simply noting the striking commonalities across the RD and participant matrices themselves (see black lines in Fig. 4), but this is also confirmed statistically ($Z = 7.52$, $p < .0001$)—and a likelihood ratio suggests that the empirical data are 56 times more likely under the transformation, relative to hypercube, model.

6. Conclusion

In this paper, we provide a novel demonstration of why relational structure even in extremely simple, stylized stimuli (let alone realistic real-world objects or scenes) can make meaningful embedding in spatial representations impossible. This limits the utility of such frameworks as methodological tools in cognitive neuroscience. And, given that humans effortlessly process and represent such relational structure, it invalidates widely used spatial schemes as a general-purpose "representational format" for complex human cognition. Of course, there may be contexts where extant techniques, from MDS through to various types of clustering, remain practically useful. Our goal is not to end the use of those methods. Rather, our goal is to help researchers understand for what kinds of stimulus properties the limitations of such approaches will be severe. There may be contexts where spatial models provide useful approximations (e.g., Herbert et al., 2020), but in order to assess adequately that utility a clear understanding of the way in which the limitations of spatial models manifest seems essential. Closer scrutiny may, for example, ultimately suggest that researchers would benefit from working with proximity matrices directly, that is, collecting empirical similarity data, rather than forcing similarities into an incompatible representational model.

7. Methods

7.1. Participants

A total of 30 Cardiff University students completed the free sorting task (mean age = 19.2, range = 18–25). An independent sample of seven participants also completed a pairwise similarity rating task. For both tasks, participants were allocated course credit for taking part and were tested individually in a cognitive testing laboratory. This study was approved by the

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Cardiff University School of Psychology local ethics committee, and all participants provided informed consent prior to taking part.

A total of 262 participants took part in the subsequent hypercube similarity rating task—that is, for the 16-item stimulus domain outlined in Fig. 3 (mean age = 28.9, range = 18–35). All participants were recruited from the participant recruitment platform Prolific (https://prolific.ac) and reimbursed for their participation at a rate of £7.50/h. Participants were between 18 and 35 years of age, UK residents, and fluent English speakers. Based on a priori power analysis (see: https://doi.org/10.17605/OSF.IO/QGFRM), a sample of $N = 146$ provides 95% power to observe a significant difference between two correlation coefficients. A larger sample of $N = 250$ was targeted to provide additional power and protect against participant attrition after data cleaning. This study was approved by the Royal Holloway Department of Psychology Research Ethics Committee.

7.2. Stimulus sets

7.2.1. Free sorting task

The stimulus set consisted of nine pairs of geometric shapes (Fig. 1). As noted in the main text and Fig. 1a, we can refer to each potential item (a shape pair) and comparison (a comparison between two shape pairs) using a letter code, where every letter corresponds to a unique feature on the shape stimulus dimension (A = square; B = triangle; C = circle). For example, the seed item in the free-sorting task (see "Tasks and procedure," below) comprised a square to the left of a triangle, and can be represented by the code *AB*. Likewise, the comparison between the "seed" item and the two circles can be labeled *AB/CC* (see Fig. 1a). As there was a single category reference (the category seed), the only relevant comparisons for modeling the data were those between task items and the category seed, that is, not all pairwise comparisons are carried out for this specific task.

7.2.2. "Hypercube" similarity rating task

For the second similarity ratings task, we created a new stimulus set that was a variation of the items presented in our free-sorting task described above. Each item in the set could be described as a 4-bit binary vector that represents the following stimulus aspects: triangle/square and light/dark for the left-hand component, and triangle/square and light/dark for the right-hand component. Exhausting all possible combinations yields a stimulus set of 16 item pairs (Fig. 3a).

7.3. Tasks and procedure

7.3.1. Free sorting task

In this task, participants were given initially the category "seed" (item *AB*, see "Stimulus set," above). A "seed" item is a single stimulus that acts as a category reference (Regehr & Brooks, 1995). After studying this "seed," the eight comparison items were presented and the whole array was available (see Fig. 1a). Participants were instructed to use the "seed" as a basis for their judgments and to choose items that they felt belonged with it. The seed method

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

was applied because engaging participants in all possible pairwise comparisons between individual items would have resulted in several "repeats," that is, equivalent comparisons with matching predictions. Having participants compare the seed with all other task items ensured that each comparison was unique and could be easily mapped onto the model predictions (see "Model predictions" below). To make the task more engaging, we presented each item as "runes" on an alien spacecraft and created a back-story in which the participants were required to classify spacecrafts as hostile or friendly. The back-story was presented on an A4 sheet in front of the participants. At the bottom of the sheet was an envelope that concealed category seed. The spaceships were 8 cm × 5 cm with the pair located in the center of each. Each shape was 0.7 cm × 0.7 cm and was separated by a horizontal distance of 0.2 cm. Stimuli were presented on a laminated card that was 12 cm × 6 cm. Participants were tested individually and were first instructed to read the back-story before inspecting the category seed. After reading the story, they were asked to open the envelope and reveal the category seed. After inspecting the seed item, they were handed the nine task items and asked to organize them as described on the instruction sheet.

7.3.2. "Free sorting" similarity rating task

To derive proximity data for MDS in an independent sample of participants (see "Model predictions" below), we used a similarity rating task involving all pairwise comparisons between items used in the free sorting task. The task was presented on a 19'' LCD monitor using Inquisit software (Millisecond Software, Seattle, WA). There were 81 trials in total (9×9) and each trial involved a comparison between two items from the stimulus set, one presented on the left of the display, and one on the right. Each possible comparison was presented twice, counterbalancing for the left-right position of items. As in the classification task, each item was a pair of geometric shapes (Fig. 1a) superimposed on a line drawing of a spacecraft. Participants could take as long as they wanted on each trial. The task took approximately 7 min to complete.

7.3.3. "Hypercube" similarity rating task

To derive proximity data for the larger 16-item set (Fig. 3a), we conducted an online similarity rating task using the online platform, Gorilla (www.gorilla.sc). To ensure instructions and stimuli could be properly viewed, only participants with access to laptops and desktop computers were eligible for the study. As participants were required to compare all items/pairs, there were 136 comparisons overall (120 unique comparisons + 16 self-comparisons). Items were presented on the left and right of the display, with the assignment of item to location randomized on a trial-wise basis. Participants were asked to judge the similarity on a 7-point scale from 1 (low similarity) to 7 (high similarity). Stimuli remained on the screen until participants provided a response by selecting a rating with their mouse. To control for the intrinsic aspects of the stimuli (e.g., the similarity between particular colors in color space), we created two stimulus sets (A and B) that had the same dimensional structure but had different color assignments. The task took approximately 12 min to complete. Task stimuli can be found at: https://osf.io/346pv/

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7.4. Model predictions

7.4.1. Spatial model

According to the spatial model, an object's position within the space is determined by its values along a set of relevant psychological dimensions. Distance between exemplars within this psychological metric space, therefore, can be considered to reflect similarity, whereby nearby items are highly similar, and distant items are highly dissimilar. Distance in geometric models is defined as:

$$d_{ij} = [S \left| x_{im} - x_{jm} \right| r]^{1/r}, \tag{1}$$

where d_{ij} is the distance between objects i and j, x_{im} is the value of object i on dimension m, and the parameter r refers to the spatial metric employed (i.e., Euclidean [$r = 2$] or city-block [$r = 1$]). In addition to being a model of similarity and mental representation, the spatial model also provides a technique for summarizing and displaying similarity data. The statistical procedure of MDS (Shepard, 1957) will generate, as its output, a spatial representation of minimum dimensionality that preserves the proximities in the input data as best possible. The goodness of this fit is reflected in the "stress" of a given solution (known as Stress-1). A high "stress" value indicates that the input data has been distorted greatly in generating the output. The distances between objects in this output representation can be used to predict patterns of data in appropriate psychological models or fit-related sets of similarity data (Frost & Gati, 1989; Nosofsky, 1986). To derive spatial model predictions for the free sorting task, we applied MDS to pairwise similarity ratings from an independent set of raters using "cmdscale" in R (see above). Input data corresponded to a 9×9 similarity matrix with each cell corresponding to the mean similarity rating for that comparison (averaged across $N = 7$). Similarity predictions between each item and the seed stimulus were derived by calculating the Euclidean distances between each item and the seed.

$$d = \sqrt{(x_{item} - x_{seed})^2 + (y_{item} - y_{seed})^2}. \tag{2}$$

7.4.2. Transformational model

The similarity predictions for the transformational model (across tasks) were derived from a coding scheme described previously (Hodgetts et al., 2009). This scheme specifies three operations that can be applied to change the representation of one stimulus in this set into another. Each operation takes a reference object and modifies it, as follows: (1) "*Create*"—taking the reference object we apply this operation to create a new feature that is unique to the target pair. For example, if the base pair consists of two squares, and the target pair consists of a square and a circle, we have to create the feature "circle" as this is not present in the base pair; (2) "*Apply*"—this operation takes an object or feature that is currently available (either via its presence in the base stimulus or via the "Create" operation) and applies it to one or both of the objects in the target item. Here, applying a shape that is readily available (on a representational level) to both shapes in the target is equivalently difficult to apply it to one. Thus, once the "circle" feature has been created (using operation 1), it is considered equivalently effortful to apply it to one, or both, of the square objects; (3) "*Swap*"—this swaps

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the relative spatial position of individual features or whole objects. These transformations are applied bidirectionally to each stimulus comparison, and the maximum transformation distance is taken as the prediction for that comparison (Hahn et al., 2003; Hodgetts et al., 2009; Hodgetts & Hahn, 2012).

7.5. Statistical methods

Pearson's bivariate correlations were used to relate human perceived similarity data to similarity model predictions (see above), and *R*-squared values were derived to assess accounted variance. Likelihood ratios were also derived to compare model fits using custom code in R (see Glover & Dixon, 2004). There were no data exclusions based on performance for the free sorting task or the associated pairwise similarity task. To ensure data quality in the larger-scale hypercube similarity rating task (and derive a marker of data "noise"), we scrubbed any participants that produced a mean self-similarity value (i.e., the 16 self-comparisons depicted across the diagonal) greater than 0.5 (max = 7).

7.6. Open science practices

Task materials, data, and analysis code to run analyses/generate figures have been made available via the Open Science Framework and can be accessed at the following link: https://osf.io/346pv/. The hypotheses and analyses for the 4D hypercube experiment were preregistered and can be found here: https://osf.io/qgfrm.

Author contributions

Carl J. Hodgetts: Conceptualization, investigation, formal analysis, data curation, visualization, writing—original draft, writing—reviewing and editing. **Ulrike Hahn**: Conceptualization, investigation, formal analysis, writing—original draft, writing—reviewing and editing.

Acknowledgments

This work was supported by the Biotechnology and Biological Sciences Research Council [BB/V010549/1] and the European Commission New and Emerging Science and Technology (NEST) Scheme [516542].

References

Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, *52*(3), 347–370. https://doi.org/10.1016/j.jml.2004.12.003

Bellmund, J. L., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*(6415). eaat6766.

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–117. https://doi.org/10.1037/0033-295X.94.2.115

Bottini, R., & Doeller, C. F. (2020). Knowledge across reference frames: Cognitive maps and image spaces. *Trends in Cognitive Sciences*, *24*(8), 606–619. https://doi.org/10.1016/j.tics.2020.05.008

Buckner, C. J. (2023). *From deep learning to rational machines: What the history of philosophy can teach us about the future of artificial intelligence* (1st ed.). New York: Oxford University Press. https://doi.org/10.1093/oso/9780197653302.001.0001

Cheries, E. W., Newman, G. E., Santos, L. R., & Scholl, B. J. (2006). Units of visual individuation in rhesus macaques: Objects or unbound features? *Perception*, *35*(8), 1057–1071. https://doi.org/10.1068/p5551

Close, J., Hahn, U., Hodgetts, C. J., & Pothos, E. M. (2010). Rules and similarity in adult concept learning. In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.), *The making of human concepts* (pp. 59–65). Oxford University Press. https://doi.org/10.1093/acprof:oSo/9780199549221.003.03

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468. https://doi.org/10.1126/science.aaf0941

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, *21*(4), 449–498. https://doi.org/10.1017/S0140525x98001253

Frost, R., & Gati, I. (1989). Comparison of the geometric and the contrast models of similarity by presentation of visual stimuli to the left and the right visual fields. *Brain and Cognition*, *9*(1), 1–15. https://doi.org/10.1016/0278-2626(89)90041-9

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (1st ed., pp. 199–241). Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.011

Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review*, *11*(5), 791–806. https://doi.org/10.3758/BF03196706

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology*, *23*(2), 222–262. https://doi.org/10.1016/0010-0285(91)90010-L

Goodhill, G. J., Simmen, M. W., & Willshaw, D. J. (1995). An evaluation of the use of multidimensional scaling for understanding brain connectivity. *Philosophical Transactions of the Royal Society B*, *348*, 265–280.

Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin*, *132*(6), 920–945. https://doi.org/10.1037/0033-2909.132.6.920

Hahn, U. (2014). Similarity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(3), 271–280. https://doi.org/10.1002/wcs.1282

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, *33*(2), 289–302. https://doi.org/10.3758/BF03195318

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*, 1–32. https://doi.org/10.1016/S0

Hahn, U., Close, J., & Graf, M. (2009). Transformation direction influences shape-similarity judgments. *Psychological Science*, *20*(4), 447–454. https://doi.org/10.1111/j.1467-9280.2009.02310.x

Hebart, M. N., Zheng, C. Y., Pereira, F., & Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature human behaviour*, *4*(11), 1173–1185.

Hodgetts, C. J., Close, J. O. E., & Hahn, U. (2023). Similarity and structured representation in human and nonhuman apes. *Cognition*, *236*, 105419. https://doi.org/10.1016/j.cognition.2023.105419

Hodgetts, C. J., & Hahn, U. (2012). Similarity-based asymmetries in perceptual matching. *Acta Psychologica*, *139*(2), 291–299. https://doi.org/10.1016/j.actpsy.2011.12.003

Hodgetts, C. J., Hahn, U., & Chater, N. (2009). Transformation and alignment in similarity. *Cognition*, *113*(1), 62–79. https://doi.org/10.1016/j.cognition.2009.07.010

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, *85*(5), 445–463. https://doi.org/10.1037/0033-295X.85.5.445

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. https://doi.org/10.1037/0033-295X.99.1.22

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27. https://doi.org/10.1007/BF02289565

Larkey, L. B., & Markman, A. B. (2005). Processes of similarity judgment. *Cognitive Science*, *29*(6), 1061–1076. https://doi.org/10.1207/s15516709cog0000_30

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica*, *102*(2–3), 221–245. https://doi.org/10.1016/S0001-6918(98)00052-3

Leech, R., Mareschal, D., & Cooper, R. P. (2007). Relations as transformations: Implications for analogical reasoning. *Quarterly Journal of Experimental Psychology*, *60*(7), 897–908. https://doi.org/10.1080/17470210701288599

Li, M. M., Huang, K., & Zitnik, M. (2022). Graph representation learning in biomedicine (No. arXiv:2104.04883). arXiv. Retrieved from http://arxiv.org/abs/2104.04883

Love, B. C., & Gureckis, T. M. (2002). Modeling unsupervised learning with SUSTAIN. In *FLAIRS Conference* (pp. 163–167).

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*(4), 431–467. https://doi.org/10.1006/cogp.1993.1011

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Millière, R. (2024). Language models as models of language (No. arXiv:2408.07144). arXiv. https://doi.org/10.48550/arXiv.2408.07144

Milton, F., & Wills, A. J. (2004). The influence of stimulus properties on category construction. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(2), 407–415. https://doi.org/10.1037/0278-7393.30.2.407

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, *4*(MAR), 128. https://doi.org/10.3389/fpsyg.2013.00128

Nickel, M., & Kiela, D. (2017). Poincaré embeddings for learning hierarchical representations (No. arXiv:1705.08039). arXiv. https://doi.org/10.48550/arXiv.1705.08039

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, *23*(1), 94–140. https://doi.org/10.1016/0010-0285(91)90004-8

Odlyzko, A. M., & Sloane, N. J. A. (1979). New bounds on the number of unit spheres that can touch a unit sphere in *n* dimensions. *Journal of Combinatorial Theory, Series A*, *26*(2), 210–214. https://doi.org/10.1016/0097-3165(79)90074-8

Op De Beeck, H., Wagemans, J., & Vogels, R. (2003). Asymmetries in stimulus comparisons by monkey and man. *Current Biology*, *13*(20), 1803–1808. https://doi.org/10.1016/j.cub.2003.09.036

Regehr, G., & Brooks, L. R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 347–363. https://doi.org/10.1037/0278-7393.21.2.347

Roads, B. D., & Love, B. C. (2024). Modeling similarity and psychological space. *Annual Review of Psychology*, *75*(1), 215–240. https://doi.org/10.1146/annurev-psych-040323-115131

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*(4), 325–345. https://doi.org/10.1007/BF02288967

Sorscher, B., Ganguli, S., & Sompolinsky, H. (2022). Neural representational geometry underlies few-shot concept learning. *Proceedings of the National Academy of Sciences*, *119*(43), e2200800119. https://doi.org/10.1073/pnas.2200800119

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Stewart, N., & Brown, G. D. A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. *Journal of Mathematical Psychology*, *49*(5), 403–409. https://doi.org/10.1016/j.jmp.2005.06.001

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.4.327

Tversky, A., & Hutchinson, J. W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review*, *93*(1), 3–22. https://doi.org/10.1037/0033-295X.93.1.3

Supporting Information

Additional Supporting Information may be found online in the supporting information tab of this article:
Supplementary Information

17568765, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tops.70001 by EBMG ACCESS - ETHIOPIA, Wiley Online Library on [10/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License